



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

AN APPLICATION OF STATISTICAL METHOD

F. A. DEWEY
Bryn Mawr, Pennsylvania

Statistics, from the standpoint of the social sciences, has remained largely a field for hope and for study. Articles appear from time to time explaining some of the elementary processes. Articles also mention that the development of statistical method carries with it hope for many of the less exact sciences. One cannot escape a feeling that some great statistical formula or method is eventually to be found which will do away with all the uncertainties or doubts that beset the path of social science.

Whether or not such a formula or set of formulae is ever to appear must for various reasons remain doubtful. One obstacle is the difficulty of dealing with mathematics that involves more than a few unknowns of any complexity. Another is the difficulty of putting human phenomena and the common experiences of life into mathematical form. Neither difficulty, however, need stand in the way of the worker who is eager to see a more exact representation of fact in the field which is covered by the social sciences, whether it be sociology, economics, social technology or economy, or any other. Not only can observed facts be summed up precisely through the use of existing statistical methods, but the use of these will in itself lead to a more careful gathering of data, and to a much more careful presentation of the data after the facts are gathered.

To illustrate, it is possible to turn to a field where the data are unquestioned and where the use of a statistical summary would have made the methods of presenting the data scientifically unimpeachable and precise. In November, 1914, for example, there appeared¹ a tabulation relating the rent which Lithuanians in Chi-

¹ *American Journal of Sociology*, November, 1914, p. 297.

cago paid for apartments and the number of rooms in the apartments. The table was as follows:

TABLE V

NUMBER OF APARTMENTS FOR WHICH SPECIFIED MONTHLY RENTALS ARE PAID AND
NUMBER OF ROOMS IN APARTMENT

RENT PER MONTH	NUMBER OF ROOMS							No Report	TOTAL
	1	2	3	4	5	6	7 or More		
Less than \$ 4.00.....	1	2	3
\$ 4.00 and less than 5.00.....	1	8	1	10
5.00 and less than 6.00.....	11	4	8	23
6.00 and less than 7.00.....	3	16	15	1	35
7.00 and less than 8.00.....	3	14	23	1	41
8.00 and less than 9.00.....	8	67	1	2	78
9.00 and less than 10.00.....	12	59	3	74
10.00 and less than 11.00.....	1	109	12	5	127
11.00 and less than 12.00.....	56	12	1	69
12.00 and less than 13.00.....	68	25	8	3	104
13.00 and less than 14.00.....	1	21	36	10	1	69
14.00 and less than 15.00.....	9	13	7	1	30
15.00 and less than 30.00.....	3	9	15	27	3	1	58
30.00 and over.....	3	5	8
Rent unknown.....	2	4	2	1	9
Rent free.....	1	2	1	4
Rent for apartment includes rent for store.....	1	1	5	23	11	5	1	1	48
									790

The author of the article accompanying the table proceeds: "Before leaving Table V it remains to see whether there is any apparent relation between the rate of rent and the number of rooms obtained. Little connection is noticeable. For example, the cost of two-room apartments ranges from less than \$4.00 monthly to between \$10.00 and \$11.00. . . ."

The author nowhere notes that 19 of the 28 two-room apartments rented for between \$4.00 and \$6.00 per month, nor was any coefficient of correlation calculated. The work of calculating the coefficient of correlation is a simple task and would obviate a clumsy and not altogether accurate statement like the one made. And furthermore the table, if it had appeared with an accurate mathe-

matical coefficient showing the relations between rooms and rent, could not have appeared with foreshortened classes. Of the 790 apartments rented 12 per cent are in one of the last three classes (rent unknown, free, or including a store) or else in undefined classes ("less than \$4.00" rent; "\$15.00 and less than \$30.00," "\$30.00 and over," rent; "7 or more" rooms). The uncertainties introduced by the foreshortening account for 74 of the 130 apartments which cannot be classified.

As the data are presented, it is only possible for the person to whom the original data are inaccessible to calculate the coefficient of correlation for that part of the table which is complete; that is, to obtain it for the 660 apartments which paid at least \$4.00 rent and less than \$15.00 rent. The inclusion of the five apartments in this part of table which have "7 or more" rooms as 7 rooms does not greatly affect the coefficient of correlation. As the inclusion decreases the coefficient, they are included.

Rent per Month	fd_x	fd_x^2	Product-Moment fd_xfd_y
\$ 4.00 and less than \$ 5.00.....	— 60	360	120
5.00 and less than 6.00.....	—115	575	130
6.00 and less than 7.00.....	—140	500	84
7.00 and less than 8.00.....	—123	369	57
8.00 and less than 9.00.....	—156	312	6
9.00 and less than 10.00.....	— 74	74	9
10.00 and less than 11.00.....	—668
11.00 and less than 12.00.....	69	69	14
12.00 and less than 13.00.....	208	416	100
13.00 and less than 14.00.....	207	621	174
14.00 and less than 15.00.....	120	480	120
	604	3,836	814
	—668		
	— 64		

The summary of the work according to the product-moment (Pearson's)¹ method is as shown in the table above. The product-moments are positive, as given, and are summed for each rent

¹ Cf. Yule, *An Introduction to the Theory of Statistics*.

row, the negative products being subtracted in each row before the total is set down.

Number of Rooms	fd_y	fd_y^2
1.....	— 3	9
2.....	— 52	104
3.....	— 156	56
4.....	— 111
5.....	104	104
6.....	66	132
7 and over.....	15	45
	185	450
	— 111	
	74	

The coefficient of correlation is then

$$\frac{814 + \left(660 \frac{74}{660} \times \frac{64}{660} \right)}{\sqrt{\left(450 - \frac{(74)^2}{660} \right) \left(3836 - \frac{(64)^2}{660} \right)}}$$

or

$$\frac{821.2}{\sqrt{441.7 \times 3830}} = +0.632.$$

And the probable error of the coefficient is

$$\frac{0.67(1-0.63^2)}{\sqrt{660}} = \frac{0.67(0.3969)}{25.7} = 0.0157,$$

which makes the correct expression of the relation between the number of rooms obtained for a given rent the following:

$$r = +0.632 \pm 0.016.$$

Since the coefficient is applicable to a part only of the table, it is obvious that for a part of the flats there is some connection between the number of rooms and the rent; location, repairs, position, etc., in other words, are not the predominating factors in rent for a part of the table.

In another article, which appeared in January, 1913,¹ a tabulation relating rent and the number of rooms in apartments is from

¹ *American Journal of Sociology*, p. 522 for date given.

the scientific standpoint a greater offender than the one just quoted. Rooms are classified only as far as "6 and over," and rent to "\$15.00 and over." Sixteen and eight-tenths of the rented apartments are indefinitely classified because of this fact. Here, using the available part of the table, but counting 6 rooms and over as 6 rooms, the coefficient of correlation by the product-moment method is 0.786 ± 0.008 (1,082 apartments included).

On p. 538 of the latter article another tabulation of the same character appears. In this a complete tabulation of rooms is given, though rent stops at "\$12.00 and over." Only 11.8 per cent of all the apartments (123) are incapable of definite classification. The coefficient of correlation for those which can be definitely classified (104) is 0.463 ± 0.050 .

In the text the median rental for 6-room apartments was found to be much higher than that for the district in which the coefficient of correlation was 0.786. The author states that "it is perhaps hardly fair to make use of the median rental for 4-room apartments, since we have the facts for only 20 such apartments" (p. 536). The relation of any apartment of a given size to its rent can be expressively stated by the proper coefficient of correlation, i.e., 0.463 ± 0.050 . Inasmuch as the true coefficient probably lies somewhere between 0.413 and 0.513, comparisons with other data are obviously not such as would inspire confidence.

This is a brief illustration of one way of many in which common statistical method can be put to good use in the field of social technology or economy. The service rendered is not alone on the side of accuracy. The use of the method reacts back into the data gathered, the method of gathering it, and its presentation. The application of the coefficient enables anyone to test the value of work being done long before it is necessary to make a final summary and draw conclusions.

There is a further, though perhaps minor, consequence of the use of precise expressions. The expressions themselves gain currency and gradually acquire meaning in a world outside, which might well become aware of the possibility of accurate statement.